Null space of the stoichiometrix matrix

- Any flux vector $\mathbf{v}$ that the cell can maintain in a steady-state is a solution to the homogeneous system of equations

$$S\mathbf{v} = \mathbf{0}$$

- By definition, the set

$$\mathcal{N}(S) = \{\mathbf{u} | S\mathbf{u} = 0\}$$

  contains all valid flux vectors
- In linear algebra $\mathcal{N}(A)$ is referred to as the null space of the matrix A
- Studying the null space of the stoichiometric matrix can give us important information about the cell's capabilities

Null space of the stoichiometric matrix

The null space $\mathcal{N}(S)$ is a linear vector space, so all properties of linear vector spcaes follow, e.g:

- $\mathcal{N}(S)$ contains the zero vector, and closed under linear combination: $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{N}(S) \implies \alpha_1 \mathbf{v}_1 + \alpha \mathbf{v}_2 \in \mathcal{N}(S)$
- The null space has a basis $\{\mathbf{k_1}, \ldots, \mathbf{k}_q\}$, a set of $q \leq \min(n, r)$ linearly independent vectors, where r is the number of reactions and n is the number of metabolites.
- The choice of basis is not unique, but the number q of vector it contains is determined by the rank of S.

Null space and feasible steady state rate vectors

- The kernel $K = (\mathbf{k}_1, \ldots, \mathbf{k}_q)$ of the stoichiometric matrix formed by the above basis vectors has a row corresponding to each reaction. (Note: the term 'kernel' here has no relation to kernel methods and SVMs)
- K characterizes the feasible steady state reaction rate vectors: for each feasible flux vector $\mathbf{v}$, there is a vector $\mathbf{b} \in \mathbb{R}^q$ such that $K\mathbf{b} = \mathbf{v}$
- In other words, any steady state flux vector is a linear combination

$$b_1\mathbf{k}_1 + \cdots + b_q\mathbf{k}_q$$

of the basis vectors of $\mathcal{N}(S)$.

Applications of null space analysis

Three properties of the metabolic network can be found directly from the kernel matrix

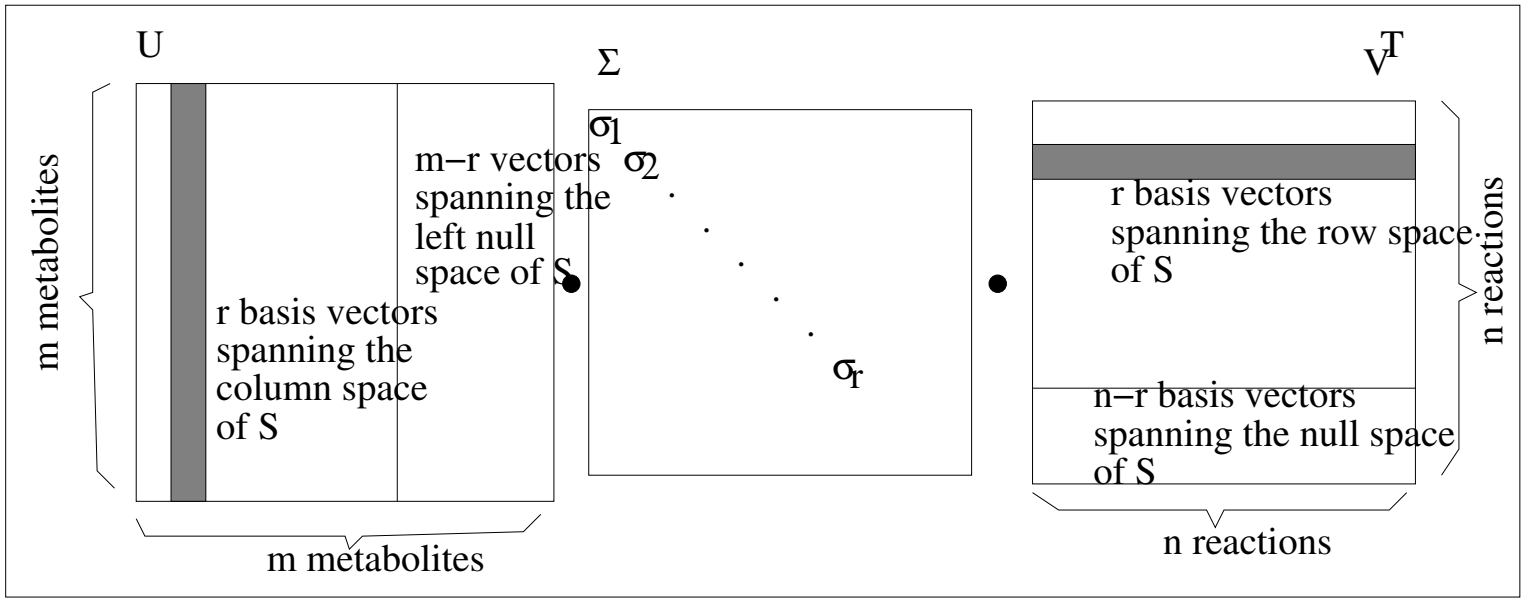
- Dead ends in metabolism (reactions that cannot carry a flus in any steady state): correspond to identically zero rows in the kernel
- Enzyme subsets (reactions that are forced to operate in lock step in any steedy state): correspond to kernel rows that are scalar multiples of each other
- Independent components (groups of reactions that can carry flux independently from reactions outside the group): block-diagonal structure in the kernel

Singular value decomposition of S

- Singular value decomposition can be used to discover a basis for the null space as well as three other fundamental subspaces of the stoichiometric matrix S
- The SVD of S is the product $S = U\Sigma V^T$, where
  - U is a $m \times m$ (m is the number of metabolites) orthonormal matrix (columns are normalized to length one $||\mathbf{u}|| = 1$, columns are orthogonal to each other $\mathbf{u}_i^T \mathbf{u}_j = 0$)
  - $\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_r)$ is $m \times n$ matrix containing the singular values σ_i on its diagonal. The rank of Σ (and S) is the number of non-zero signular values
  - V is a $n \times n$ orthonormal matrix (n is the number of reactions)

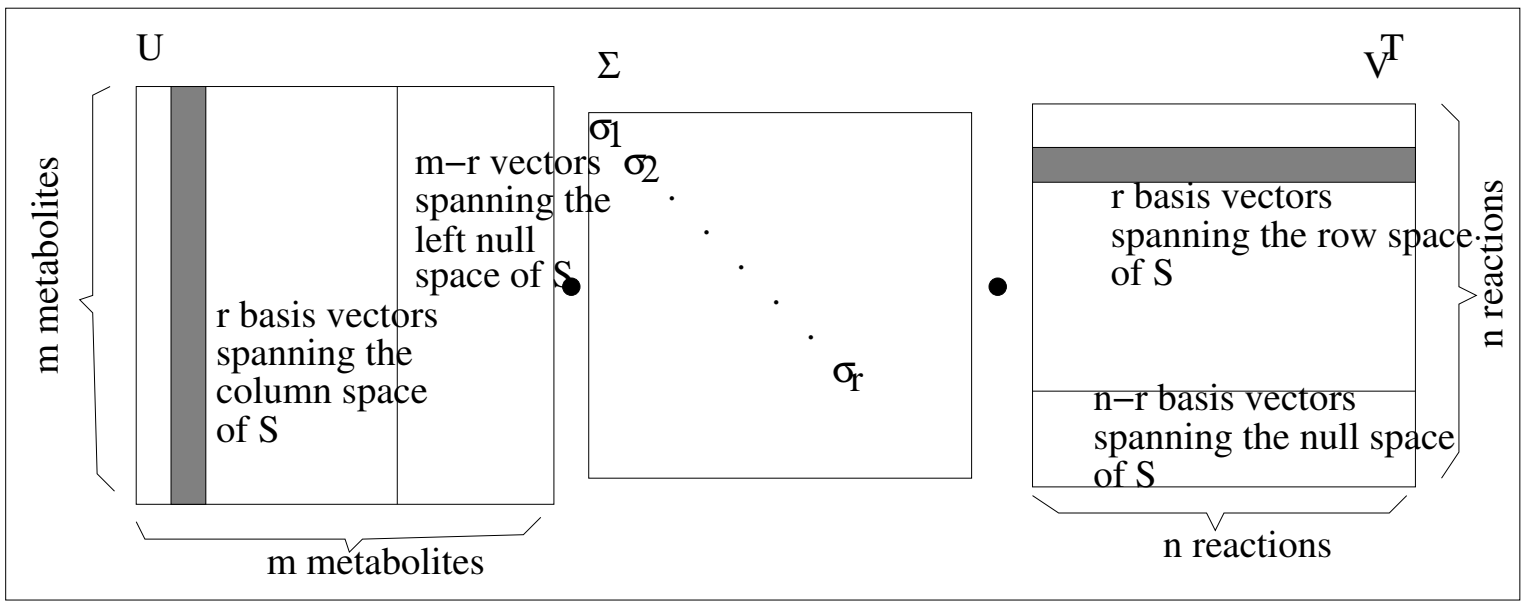
Singular value decomposition of S: matrix U

- The columns of U can be seen as as prototypical or 'eigen-' reactions
- All reaction stoichiometries in the metabolic system can be expressed as linear combinations of the eigen-reactions.
- The eigen-reactions are linearly independent, while the original reactions (columns of S) may not be (e.g. duplicate reactions)



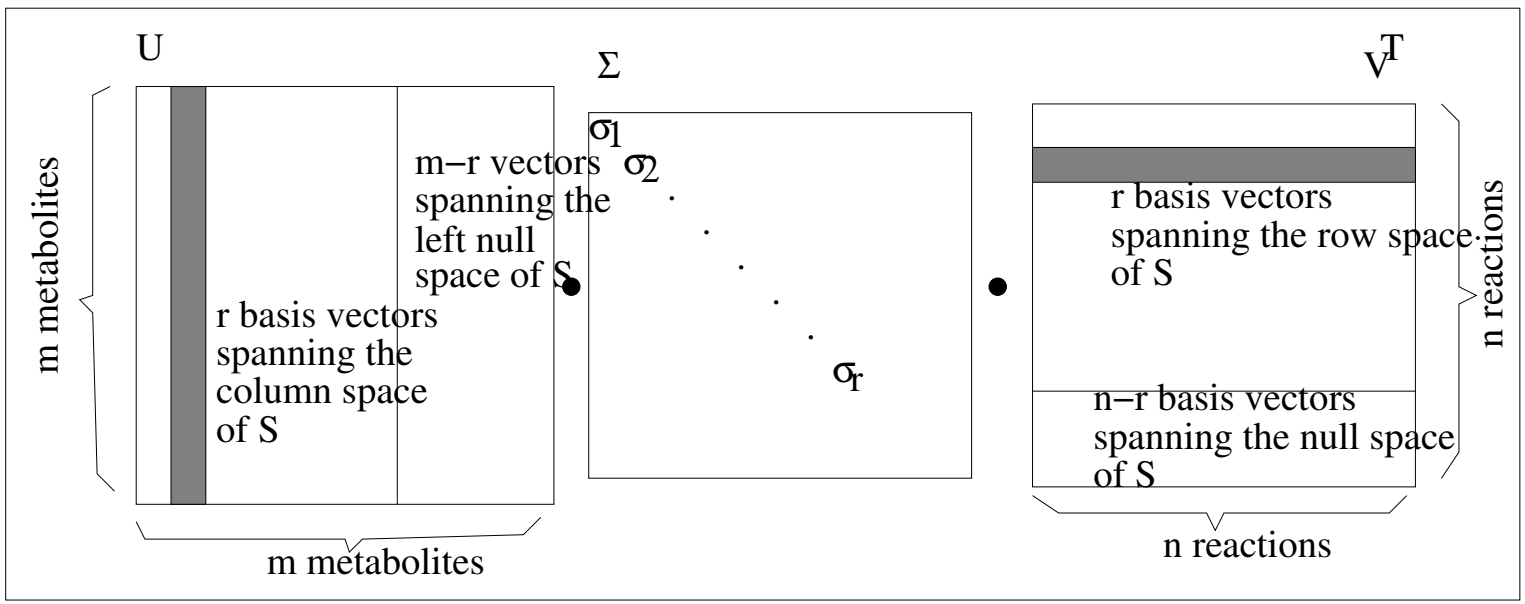
Singular value decomposition of S: matrix U

- The first r columns of S span the column space of S
- The column space contains all possible time derivatives of the concentration vector
- i.e. what kind of changes to each metabolite concentrations are possible given the network structure and the activity of the reactions



Singular value decomposition of S: matrix U

- The $m - r$ vectors $\mathbf{u}_{r+l}$ span the *left null space* of S
- Left null space of S isthe set $\{\mathbf{u}|S^T\mathbf{u} = 0\}$ (or alternatively $\mathbf{u}^T S = 0$)
- Given a vector $\mathbf{u}$ form the left null space, for any column $\mathbf{s}_j$ of S (i.e. reaction stoichiometry), the equation $\sum_i \mathbf{s}_{ij} u_i = 0$ holds

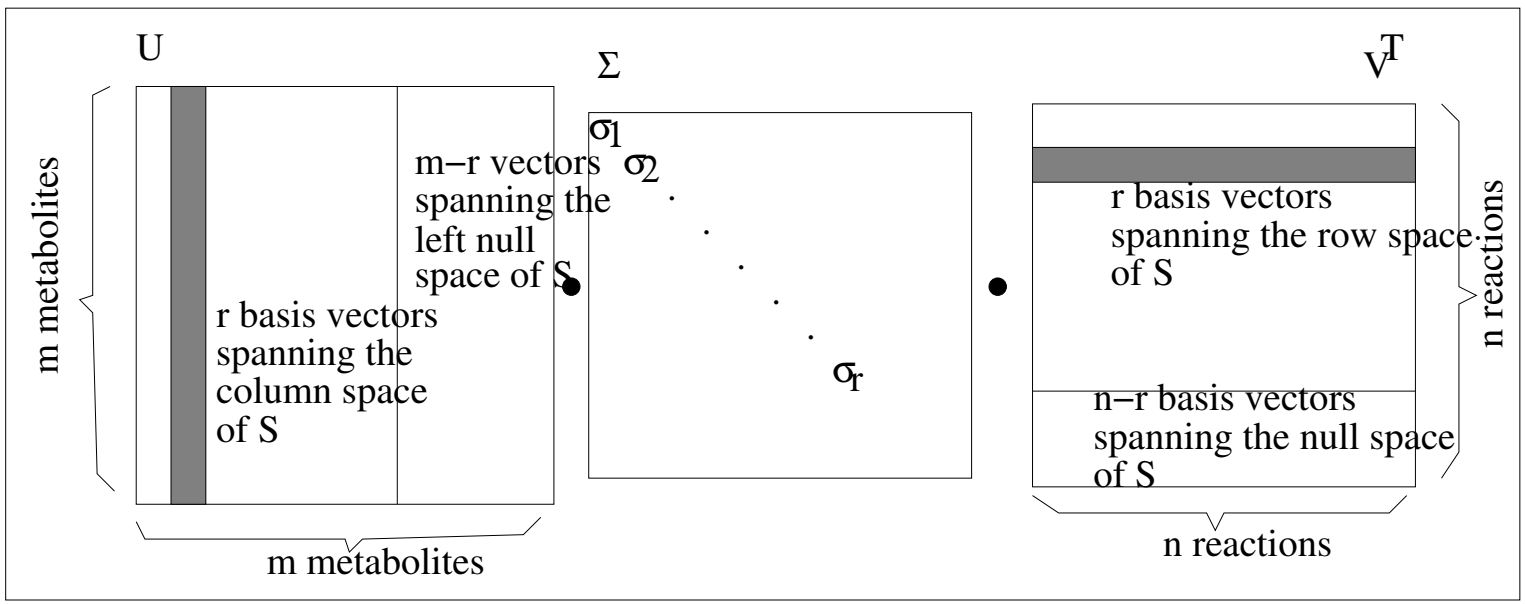


Singular value decomposition of S: matrix U

- The left null space represents metabolite conservation via the equations

$$\sum_i \mathbf{s}_{ij} u_i = 0$$

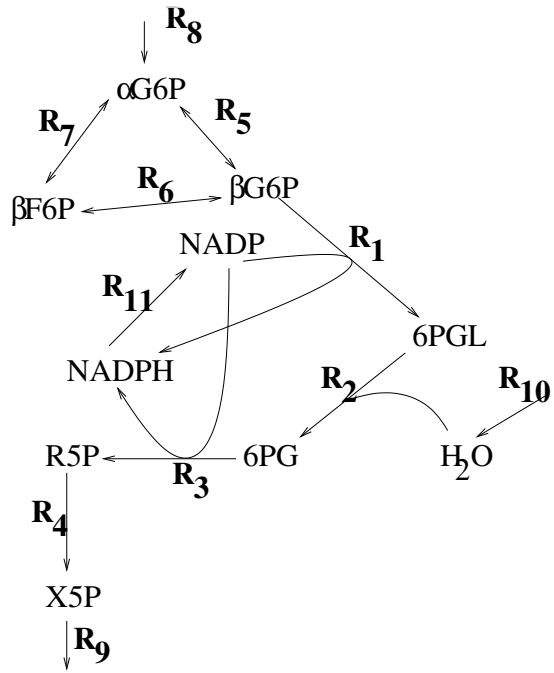
- The non-zero coefficients of the left null space vectors **u** represent pools of metabolites that remains of constant size regardless of which reactions are active and how active they are



Conservation in PPP

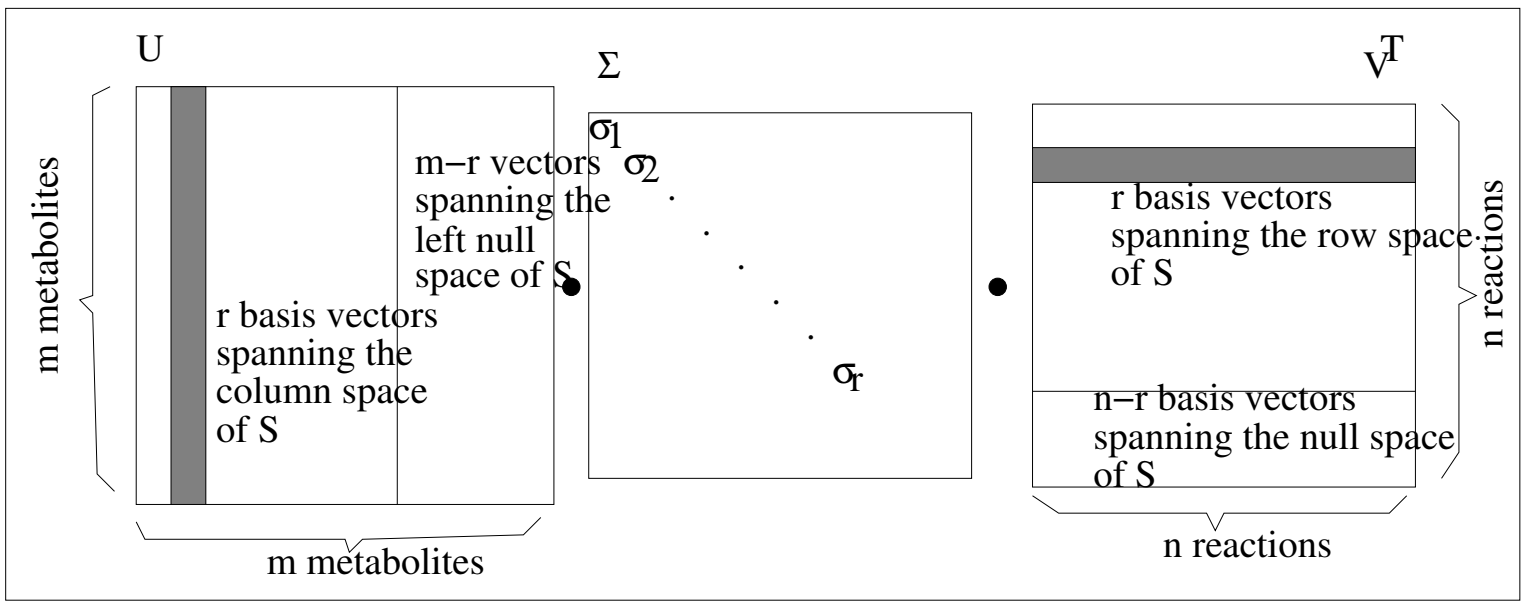
The left null space of our PPP system only contains a single vector, stating that the sum of $NADP^+$ and *NADPH* is constant in all reactions.

$$\mathbf{l}^T = \begin{matrix} \beta G6P \\ \alpha G6P \\ \beta F6P \\ 6PGL \\ 6PG \\ R5P \\ X5P \\ NADP^+ \\ NADPH \\ H_2O \end{matrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.7071 \\ 0.7071 \\ 0 \end{bmatrix}$$



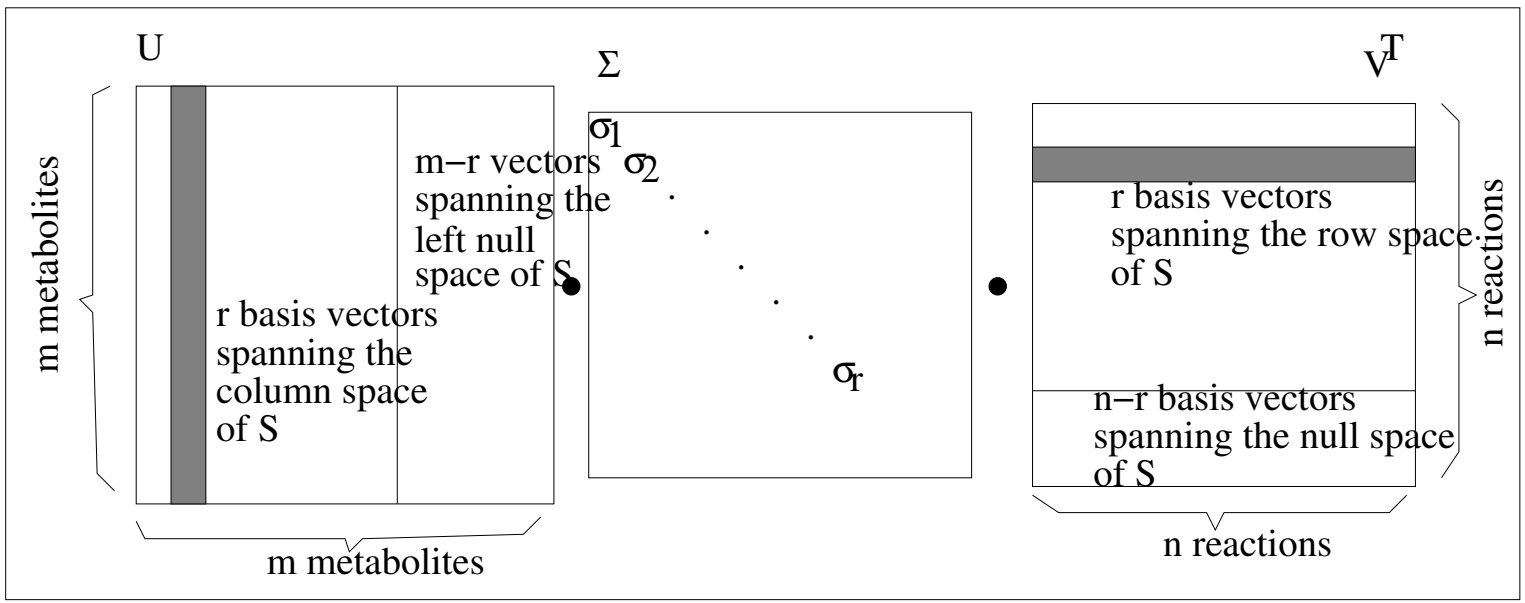
Singular value decomposition of S: matrix V

- The columns of matrix V can be seen as systems equations of prototypical 'eigen-' metabolites.
- These eigen- systems equations are linearly independent
- All systems equations of the metabolism can be expressed as their linear combinations.



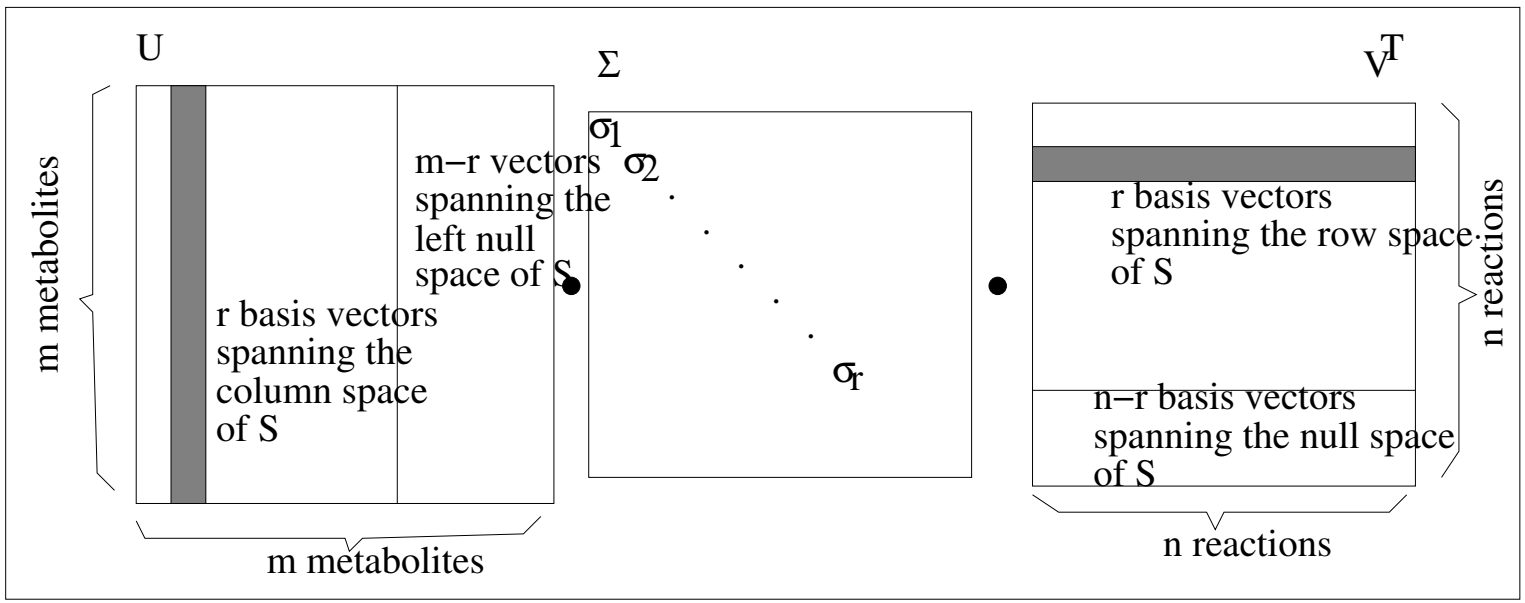
Singular value decomposition of S: matrix V

- The first r columns of V span the row space of S
- The row space contains all non-steady state reaction rate vectors that are possible for the system represented by S



Singular value decomposition of S: matrix V

- The last $n - r$ columns of V span the null space of S
- These are flux vectors that can operate in steady state, i.e. statifying $S\mathbf{v}_l = 0, l = r + 1, \ldots, n$
- These can be taken as the kernel K used to analyze steady state fluxes (this is how we obtained K previously).



SVD of PPP

MATLAB script $ppp_svd.m$ computes

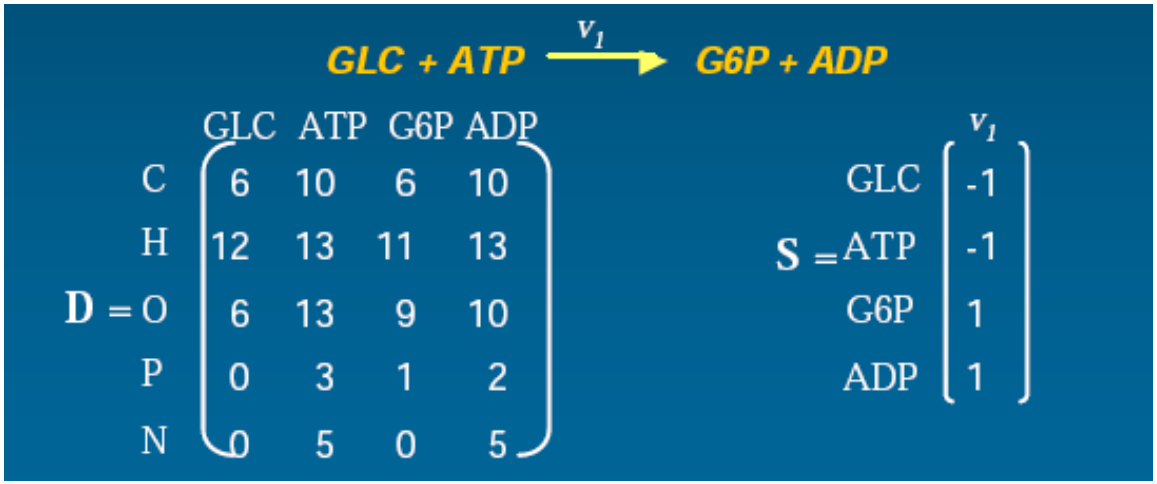
- The stoichiometric matrix S
- The singular value decomposition $S = U\Sigma V^T$
- The kernel matrix of the null space K
- The kernel matrix of the left null space K_{left}

Other conserved quantitites

- Above look at conservation of pool sizes of metabolites
- Conservation of other items can be analyzed as well:
  - Elemental balance: for each element species (C,N,O,P,...) the number of elements is conserved
  - Charge balance: total electrical charge, the total number of electrons in a reaction does not change.

Elemental balancing (1/2)

- All chemical reactions need to be elementally balanced
- The number of elements of different species (carbon, hydrogen, oxygen, ...) need to be balanced
- Let D be a matrix defining the elemental composition of the participating metabolites, and vector S denote the stoichiometric coefficients of a reaction



(picture from B Palsson course material http://gcrg.ucsd.edu/classes/)

Elemental balancing (2/2)

- Multiplication of any row of D with the stoichiometric coefficient vector should give 0
- A balance for carbons can be verified form the first row by multiplying with the stoichiometric coefficients

$$6 \cdot -1 + 10 \cdot -1 + 6 \cdot 1 + 10 \cdot 1 = 0$$

- The same calculation for hydrogen results in an error

$$12 \cdot -1 + 13 \cdot -1 + 11 \cdot 1 + 13 \cdot 1 = -1$$

- The reaction equation is not balanced, a should be corrected. The correct equation is $GLC + ATP \mapsto G6P + ADP + H$

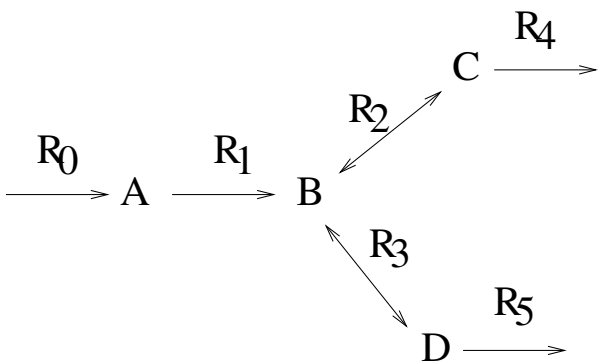
$$\mathbf{DS} = \begin{pmatrix} 6 & 10 & 6 & 10 \\ 12 & 13 & 11 & 13 \\ 6 & 13 & 9 & 10 \\ 0 & 3 & 1 & 2 \\ 0 & 5 & 0 & 5 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{matrix} \text{C} \\ \text{H} \\ \text{O} \\ \text{P} \\ \text{N} \end{matrix}$$

Basis steady state flux modes from SVD

- A basis for the null space is thus obtained by picking the $n - r$ last columns of V from the SVD of S:

$$K = [v_{r+1}, \ldots, v_n]$$

- In MATLAB, the same operation is performed directly by the command *null*(S).
- Let us examine the following simple system

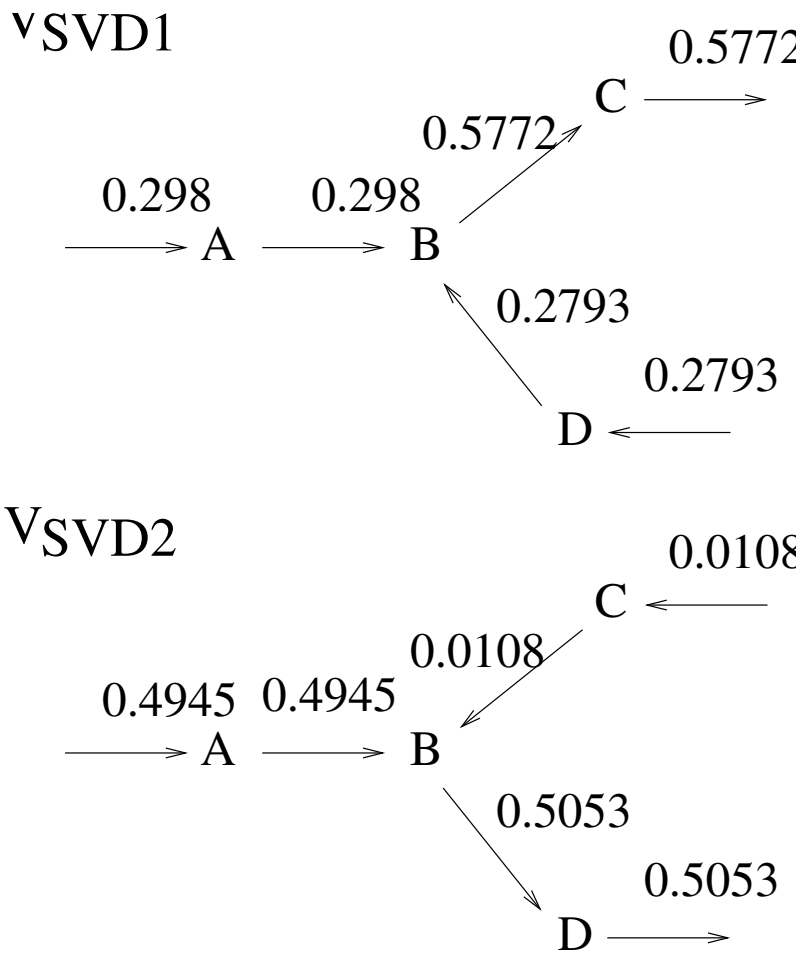


$$S = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$

Basis steady state flux modes from SVD

- The two flux modes given by SVD for our example system
- All steady state flux vectors can be expressed as linear combinations of these two flux modes

$$K = \begin{bmatrix} 0.2980 & 0.4945 \\ 0.2980 & 0.4945 \\ 0.5772 & -0.0108 \\ -0.2793 & 0.5053 \\ 0.5772 & -0.0108 \\ -0.2793 & 0.5053 \end{bmatrix}$$

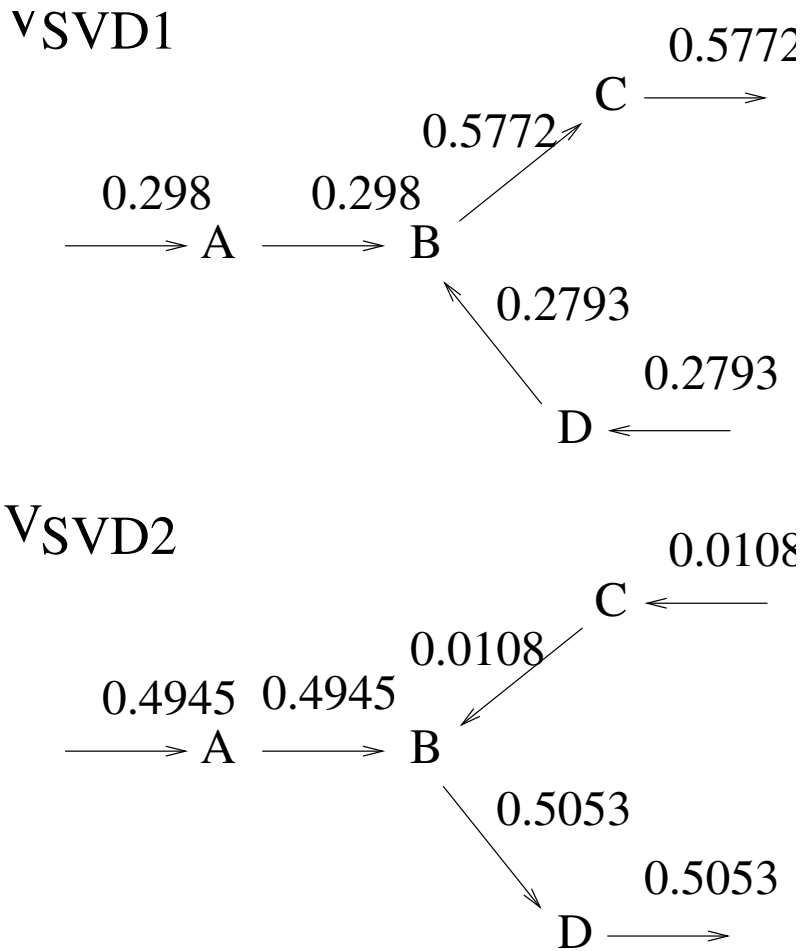


Basis steady state flux modes from SVD

The kernel matrix obtained from SVD suffers from two shortcomings, illustrated by our small example system

- Reaction reversibility constraints are violated: in v_{svd1}, $R5$ operates in wrong direction, in v_{svd2}, R_4 operates in wrong direction
- All reactions are active in both flux modes, which makes visual interpretation impossible for all but very small systems

- The flux values are all non-integral



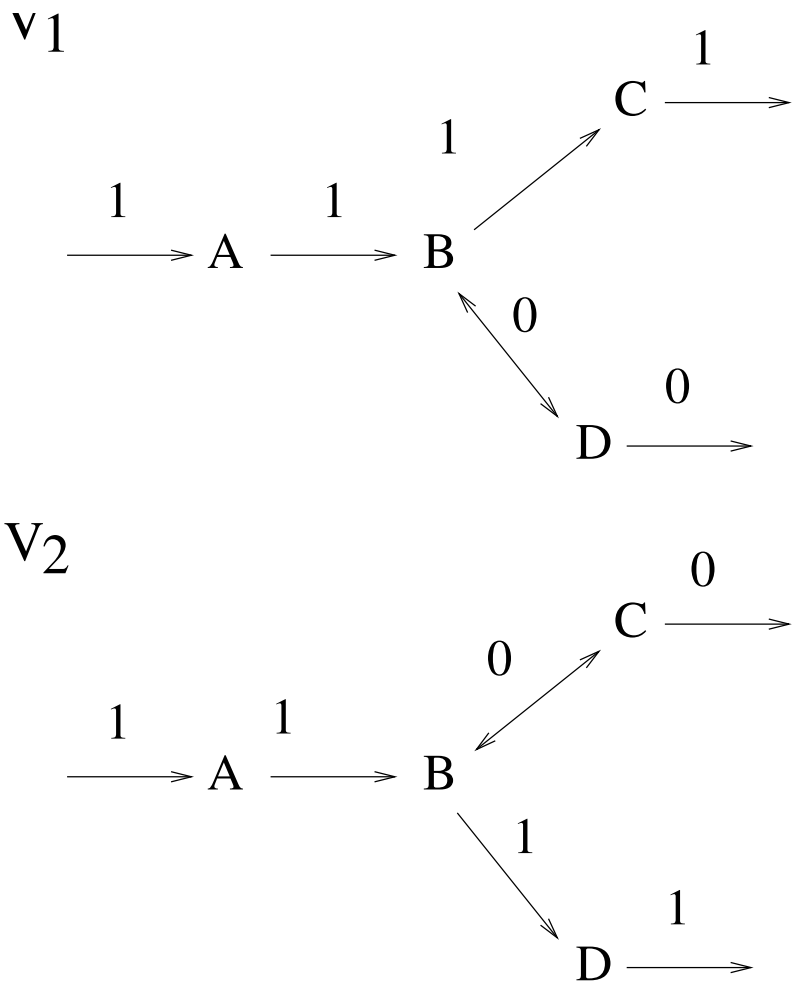
Choice of basis

- SVD is only one of the many ways that a basis for the null space can be defined.
- The root cause for hardness of interpretation is the orthonormality of matrix V in SVD $S = U\Sigma V^T$
  - The basis vectors are orthogonal: $v_{svd1}^T v_{svd2} = 0$
  - The basis vectors have unit length $||v_{svd1}|| = ||v_{svd1}|| = 1$
- Neither criteria has direct biological relevance!

Biologically meaningful pathways

- From our example system, it is easy to find flux vectors that are more meaningful than those given by SVD
- Both pathways on the right statisfy the steady state requirement
- Both pathways obey the sign restrictions of the system
- One can easily verify (by solving b form the equation $Kb = v$) that they are linear combinations of the flux modes given by SVD,

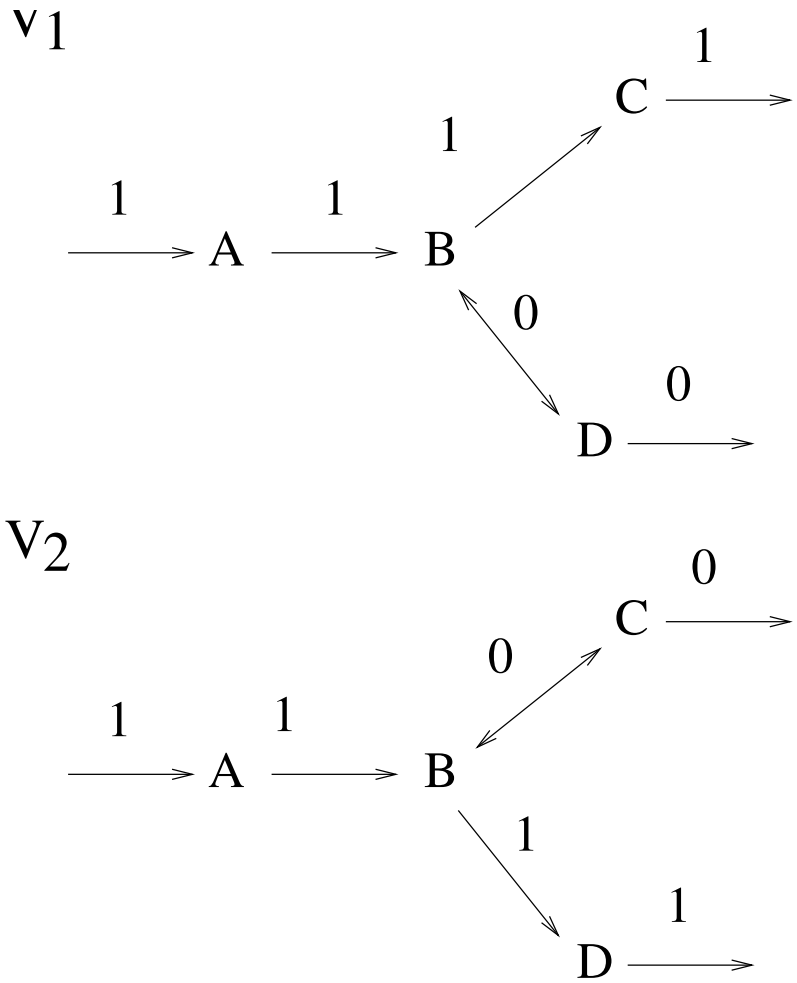
e.g. $v_1 = 0.0373 v_{svd1} + 1.997 v_{svd2}$



Elementary flux modes

The two pathways are examples of elementary flux modes
The study of elementary flux modes (EFM) and concerns decomposing the metabolic network into components that

- can operate independently from the rest of the metabolism, in a steady state,
- any steady state can be described as a combination of such components.



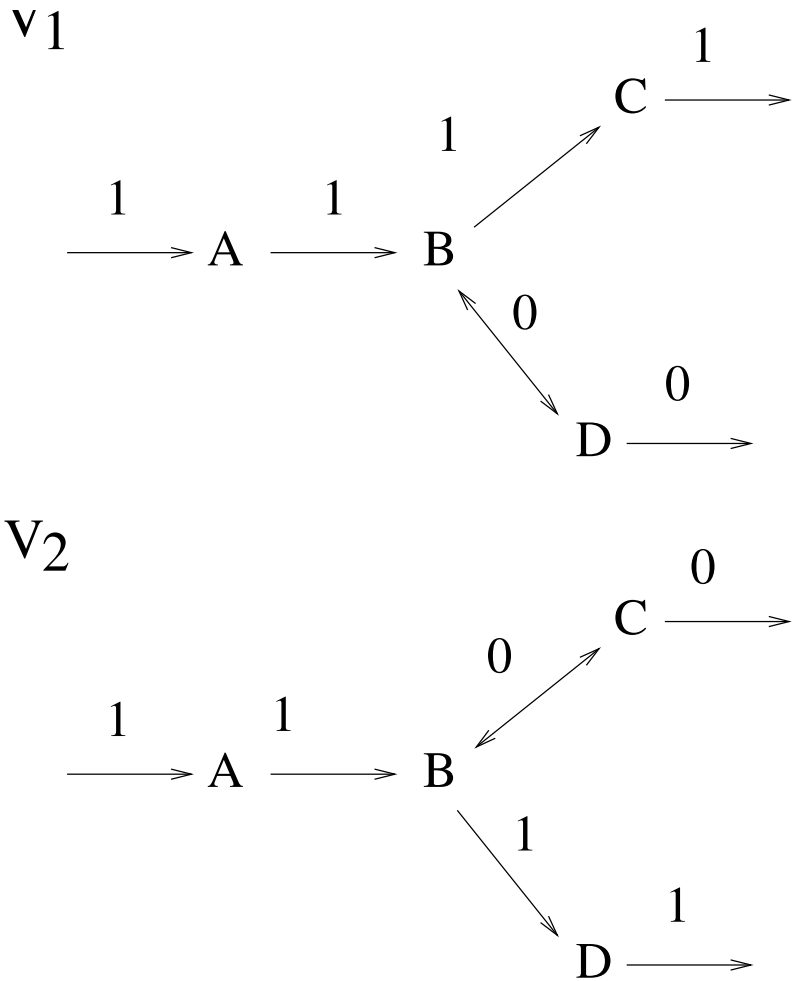
Representing EFMs

- Elementary flux modes are given as reaction rate vectors

$$\mathbf{e} = (e_1, \ldots, e_n),$$

- EFMs typically consists of many zeroes, so they represent pathways in the network given by the non-zero components

$$P(\mathbf{e}) = \{j | e_j \neq 0\}$$



Properties of elementary flux modes

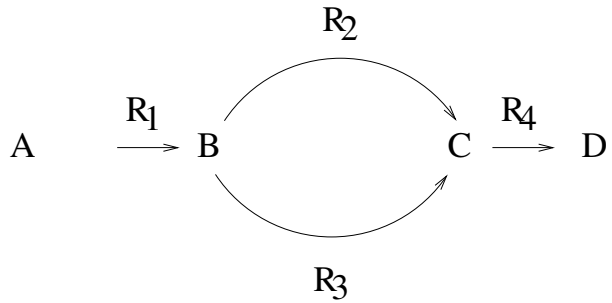
The following properties are statisfied by EFMs:

- (Quasi-) Steady state
- Thermodynamical feasibility. Irreversible reactions need to proceed in the correct direction. Formally, one requires $e_j \geq 0$ and that the stoichiometric coefficients s_{ij} are written with the sign that is consistent with the direction
- Non-decomposability. One cannot remove a reaction from an EFM and still obtain a reaction rate vector that is feasible in steady state. That is, if $\mathbf{e}$ is an EFM there is no vector $\mathbf{v}$ that satisfies the above and $P(\mathbf{v}) \subset P(\mathbf{e})$

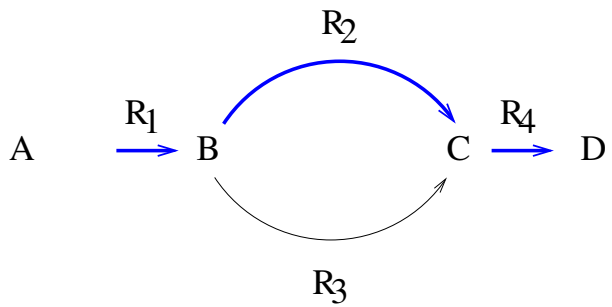
These properties define EFMs upto a scaling factor: if $\mathbf{e}$ is an EFM $\alpha\mathbf{e}, \alpha > 0$ is also an EFM.

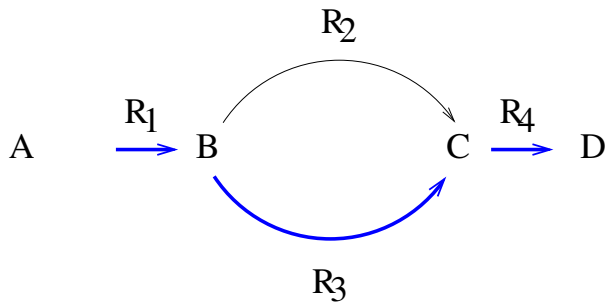
Example

Metabolic system:

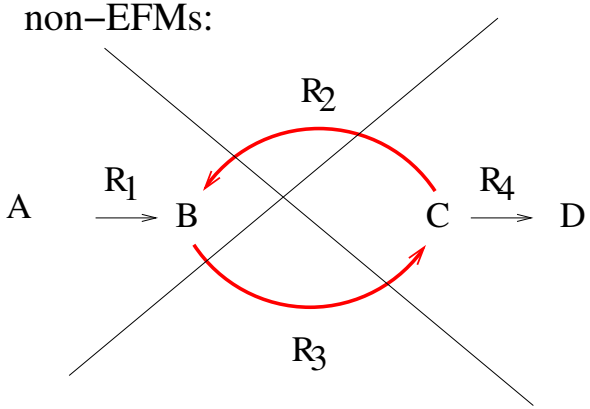


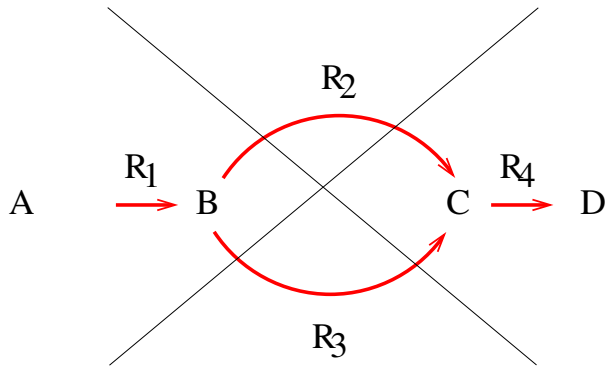
EFMs:





non−EFMs:





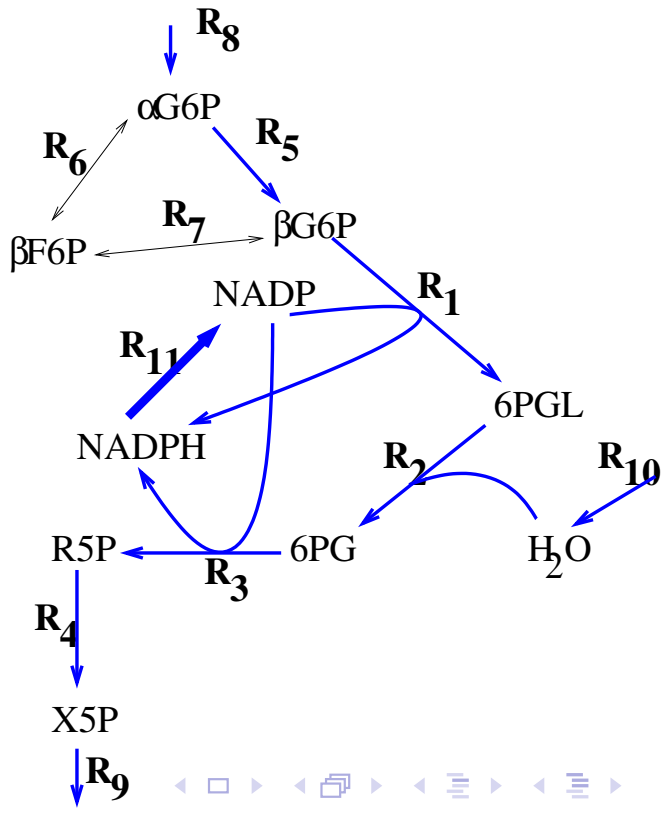
EFMs and steady state fluxes

- Any steady state flux vector $\mathbf{v}$ can be represented as a non-negative combination of the elementary flux modes: $\mathbf{v} = \sum_j \alpha_j \mathbf{e}_j$, where $\alpha_j \geq 0$.
- However, the representation is not unique: one can often find several coefficient sets α that satisfy the above.
- Thus, a direct composition of a flux vector into the underlying EFPs is typically not possible. However, the *spectrum* of potential contributions can be analysed

EFMs of PPP

- One of the elementary flux modes of our PPP system is given below
- It consist of a linear pathway through the system, exluding reactions R_6 and R_7
- Reaction R_{11} needs to operate with twice the rate of the others

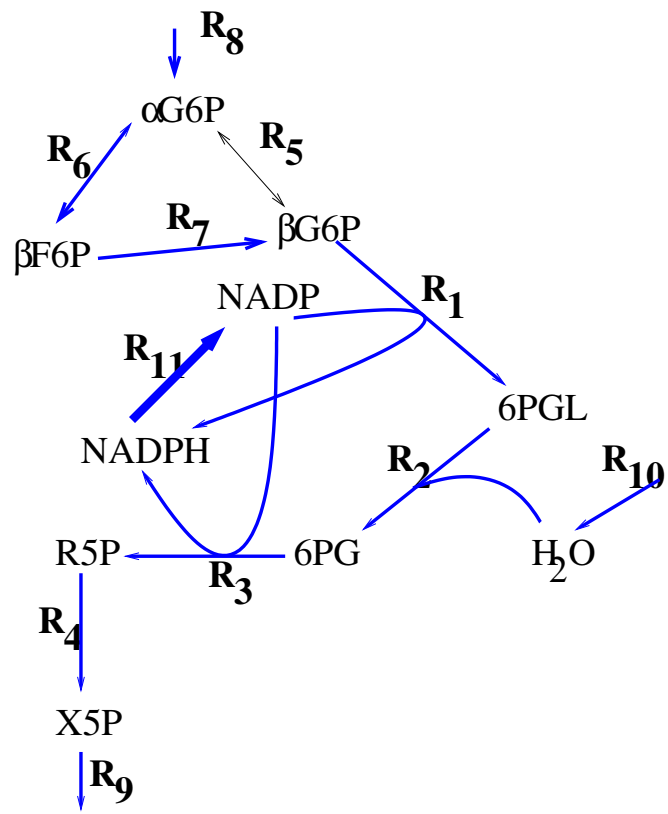
$$efm_1 = \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \\ R_7 \\ R_8 \\ R_9 \\ R_{10} \\ R_{11} \end{matrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 2 \end{bmatrix}$$



EFMs of PPP

- Another elementary flux mode of our PPP system
- Similar linear pathway through the system, but exluding reactions R_5 and using R_7 in reverse direction
- Again, reaction R_{11} needs to operate with twice the rate of the others

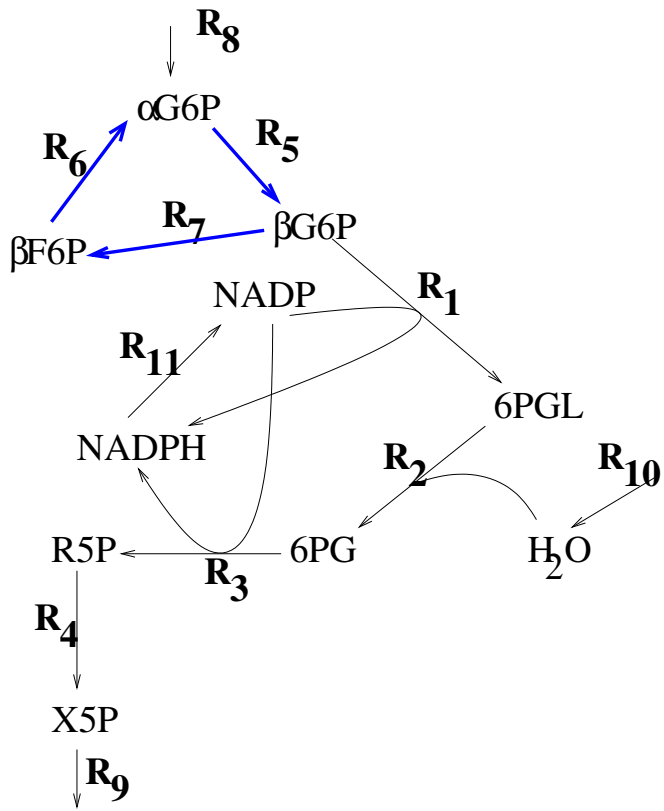
$$efm_2 = \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \\ R_7 \\ R_8 \\ R_9 \\ R_{10} \\ R_{11} \end{matrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ -1 \\ 1 \\ 1 \\ 1 \\ 2 \end{bmatrix}$$



EFMs of PPP

- Third elementary flux mode contains only the small cycle composed of R_5, R_7 and R_6. R_6 is used in reverse direction
- A yet another EFM would be obtained by reversing all the reactions in this cycle

$$efm_3 = \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \\ R_7 \\ R_8 \\ R_9 \\ R_{10} \\ R_{11} \end{matrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$



Building the kernel from EFMs

- In general there are more elementary flux modes than the dimension of the null space
- Thus a linearly independent subset of elementary flux modes suffices to span the null space
- In our PPP system, any two of the three EFMs together is linearly independent, and can thus be taken as the representative vectors

$$EFM = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \\ R_7 \\ R_8 \\ R_9 \\ R_{10} \\ R_{11} \end{array} \begin{bmatrix} \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 2 & 2 \end{bmatrix}$$

Software for finding EFMs

- From small systems it is relatively easy to find the EFMs by manual inspection
- For larger systems this becomes impossible, as the number of EFMs grows easily very large
- Computational methods have been devised for finding the EFMs by Heinrich & Schuster, 1994 and Urbanczik and Wagner, 2005
- Implemented in MetaTool package

Extreme pathways

- Extreme pathways (EP) are an alternative formalism to EFMs for analyzing the steady state flux space
- Extreme pathways differ from EFMs in two ways
  - The EPs are always non-negative $\mathbf{v} \geq 0$. Bi-directional reactions need to be represented as separate forward and backward reactions.
  - In EPs the maximum rates of the reactions are also considered $0 \leq v_i \leq v_{i,max}$

Extreme pathways

- All steady state flux vectors can be expressed as convex combinations of extreme pathways $\mathbf{p}_i$: $\mathbf{v} = \sum_i \alpha_i \mathbf{p}_i, 0 \leq \alpha_i$
- Geometrically, the extreme pathways form a high-dimensional polyhedron enclosing all legal steady state fluxes
- Flux balance analysis uses this polyhedron as the feasible set of fluxes where the flux vector optimizing the objective (e.g. biomass growth) needs to reside